Eur Respir J 2000; 15: 819–820
Printed in UK – all rights reserved

European Respiratory Journal
ISSN 0903-1936

EDITORIAL

The problem of performing adequately sized randomized trials to demonstrate small survival benefits

M. Paesmans, J-P. Sculier

Two decades of clinical research have resulted in survival improvements for advanced non-small cell lung cancer. These results have been established by multiple randomized trials completed by meta-analyses such as by the review conducted by the Non-Small Cell Lung Cancer Collaborative Group [1]. Chemotherapy, combined with radiotherapy in locally advanced disease, has indeed been shown to reduce the risk of death by 10% (reduction corresponding to a 3% absolute survival benefit at 2 yrs). Compared to best supportive care alone, mainly in the metastatic stage of the disease, chemotherapy also improves survival with a reduction in mortality of ~25%, corresponding to an increase in the median survival of 1–2 months and a 10% absolute benefit at 1 yr.

To demonstrate such a small survival benefit, the researchers should ideally conduct large size randomized trials, reaching sufficient power to detect a realistic therapeutic improvement, that will provide reliable conclusions.

One possibility is to conduct a trial with a fixed, *a priori* planned, sample size. This means, in the case of survival as the primary end-point, that the number of events required at the time of the analysis determining the study′s power, will be calculated prior to the trial activation and specified in the protocol. This approach will lead to a convincing conclusion with a narrow confidence interval for the true treatment effect. The approach will, however, cause the clinical research group to be confronted by the difficulty of successfully recruiting the required number of patients in a reasonable time period, in order to maintain enthusiasm for the trial among the participating investigators and relevance for the scientific question addressed by the trial.

An attractive alternative, for the researchers and for the patients, is to choose a design allowing for interim analyses with the hope of detecting an earlier difference, if there is one, between treatment arms and to stop the trial′s accrual prematurely. However, multiple examinations of the data inflate the probability of falsely detecting a difference (the real p-value is greater than each nominal p-value calculated at each analysis) and may bias the treatment effect estimate. Stopping boundaries, like those proposed by POCOCK [2] (using a constant nominal level at each interim analysis) or by O′BRIEN and FLEMING [3] (using nominal p-values increasing with the amount of information already collected in the trial and therefore very conservative at the beginning of a trial) are constructed to maintain the overall probability of falsely concluding a therapeutic difference. On the other hand, methods also providing unbiased treatment effect estimates for some classes of boundaries have been proposed by WHITEHEAD [4]. The boundaries are constructed using functions determining the part, already used, of the overall probability to falsely reject the hypothesis of no difference between treatment arms, called alpha-spending functions. Whichever design is adopted, it should be kept in mind that the interim examinations of the data should not be data driven. There is therefore a need to plan their number and timing as well as to fix the format of the stopping boundaries, although the decision to stop or to continue a trial is certainly a more complex issue than just the determination of a p-value [5]. In such a context, the trial′s progress should ideally be monitored by an independent data monitoring committee who should advise the investigators on whether the trial should continue or not [6]. When survival data are analysed, there is also a potential danger of drawing conclusions from short-term follow-up data, as illustrated by SYLVESTER *et al.* [7] in a practical example of a breast cancer trial.

If the monitoring policy is not carefully planned, scepticism about the trial′s conclusions might occur in the scientific community and the trial will loose the impact it could have had on the clinical practice. This happened to the well known Cancer and Leukemia Group B trial published by DILLMAN *et al.* [8] in 1990 that tested the efficacy of the addition of two cycles of induction chemotherapy prior to chest radiation in stage III non-small cell lung cancer. This randomized study was designed with a fixed sample size but was closed prematurely following estimation of a treatment effect that was only statistically significant after covariates adjustment and observation of one quarter of the events initially required [9]. The transformation of the design to a sequential one (using the boundaries of O′BRIEN and FLEMING [3]), after activation, was primarily due to an accrual rate that was much lower than expected (five patients per month instead of the expected 20 per month). The credibility of the trial results was indeed questioned, namely by SOUHAMI *et al*. [10] in a letter to the Editor of the New England Journal of Medicine.

In this issue of the *European Respiratory Journal*, a very interesting paper by DONALDSON *et al*. [11] provides a re-analysis of the Cancer and Leukemia Group B trial, with long-term follow-up data. DILLMAN *et al.* [8] used two group sequential procedures providing unbiased treatment effect estimates on simulation data (under the conservative hypothesis of no treatment effect) to increase the sample size up to the initially required one. These re-analyses confirm that the trial was closed too early, due to

Correspondence: J-P. Sculier, Institut Jules Bordet, Rue Héger-Bordet 1, B-1000 Brussels, Belgium Fax: 32 25343756.

a lack of feasibility, without sufficient statistical evidence of the superiority of the combined arm. They also confirmed that the positive result claimed by Cancer and Leukemia Group B is not reversed, either by the longer follow-up reached for the registered patients, or by the calculation of unbiased treatment effects, or by the use of conservative simulation data. However convincing these updated results may be, they come 10 yrs after the initial publication and should encourage investigators to concentrate their efforts on the participation in collaborative research, allowing the conduct of adequately sized clinical trials.

References

1. Non-Small Cell Lung Cancer Chemotherapy Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ* 1995; 311: 899–909.
2. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; 64: 191–199.
3. O′Brien TC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; 35: 549–556.
4. Whitehead J. The design and analysis of sequential clinical trials. 2nd Edn. Chichester, West Sussex, UK, Ellis Horwood, 1991.
5. Simon R. Some practical aspects of the interim monitoring of clinical trials. *Stat Med* 1994; 13: 1401–1409.
6. Pocock SJ. Statistical and ethical issues in monitoring clinical trials. *Stat Med* 1993; 12: 1459–1469.
7. Sylvester R, Bartelink H, Rubens R. A reversal of fortune: practical problems in the monitoring and interpretation of an EORTC breast cancer trial. *Stat Med* 1994; 13: 1329–1335.
8. Dillman RO, Seagren SL, Propert K, *et al*. A randomized trial of induction chemotherapy plus high-dose radiation alone in stage III non-small cell lung cancer. *N Engl J Med* 1990; 323: 940–945.
9. George SL, Chengchang L, Berry DA, Green MR. Stopping a clinical trial early: frequentist and bayesian approaches applied to a CALGB trial in non-small cell lung cancer. *Stat Med* 1994; 13: 1313–1328.
10. Souhami RL, Spiro SG, Cullen M. Chemotherapy and radiation therapy as compared with radiation therapy in stage III non-small cell lung cancer (letter). *N Engl J Med* 1991; 324: 1136–1137.
11. Donaldson N, Dillman RO, Wallace J, Ortiz-Hurtado A. Sequential re-analysis of a phase III clinical trial in non-small cell lung cancer. *Eur Respir J* 2000; 15: 821–827.